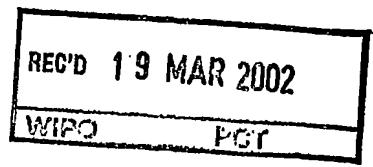




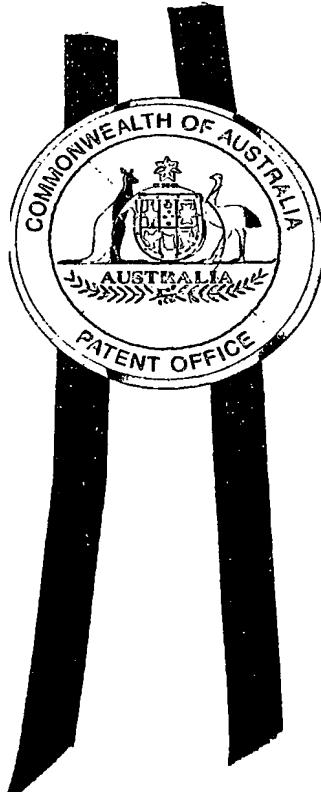
PCT/AU02/00184



Patent Office
Canberra

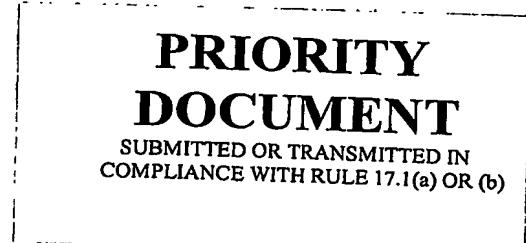
I, JONNE YABSLEY, TEAM LEADER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PR 3295 for a patent by WORLDLINGO, INC filed on 22 February 2001.

BEST AVAILABLE COPY



WITNESS my hand this
Eleventh day of March 2002

JONNE YABSLEY
TEAM LEADER EXAMINATION
SUPPORT AND SALES



AUSTRALIA

Patents Act 1990

PROVISIONAL SPECIFICATION

Invention Title: "TRANSLATION INFORMATION SEGMENT"

The invention is described in the following statement:

TRANSLATION INFORMATION SEGMENT

This invention relates generally to the field of translation, and more particularly to a translation information segment that facilitates
5 seamless translation of communications.

The invention also relates to a method and apparatus for using the translation information segment to provide seamless translation of a communication in a network environment.

10 BACKGROUND TO THE INVENTION

Machine translation of communication from one language to another is breaking down the communication barrier between individuals and businesses. Over the past twenty years there have been steady improvements in the quality of machine translation.

15 Various techniques have been developed that translate by phrase rather than word by word. Other techniques use dictionaries or translation memories to translate whole sentences. As a result the grammar of translated communications has improved and hence the readability. Some of the best translation programs are approaching
20 the quality of human translation for common languages and for specific purposes.

Although the technical ability of machine translation software has improved dramatically, the usability has improved very little. In order to translate a document, email or other communication, it is
25 generally necessary to access a translation site and run a translation program. Parameters for the program, such as source and destination language, preferred dictionary, special words, etc, must be input by the user.

30 In our co-pending United States patent application number 09/676690 we describe a one-click translation system that avoids

much of the user input that has been necessary to obtain a translation of a communication. The one-click translation system comprises a one-click translation component and a translation manager that combine to provide an almost seamless translation

5 once a user clicks the one-click component. The one-click translation system does not address the quality of the translation.

Although the one-click system is a significant advance over the prior art, it still requires some action by the receiver of the communication. For translation of communications to be universally

10 accepted, it must be completely seamless. A system is required that automatically delivers a communication in the preferred language of the recipient. The system must also deliver a better quality translation than is presently available.

Some recent technologies approach, but fail to achieve this ideal. For example, United States patent number 6161082 assigned to AT&T Corp describes a network based language translation system that aims to improve machine translation by utilizing the processing power of a network to perform the translation rather than a local machine. However, this patent fails at clarifying how the

20 detection of the involved languages is done. It only mentions that the source and target language can be detected from the communication between the two parties without indicating how this is achieved. The AT&T approach does not provide any intrinsic improvement in the quality of the translation, improvement is only achieved by increased

25 processing power available in a networked environment.

United States patent number 5548508, assigned to Fujitsu Limited, aims to improve the quality of a machine translation by embedding tags within a document that include contextual information. For example, a <TITLE> ... </TITLE> tag indicates that

30 the words are the title and should be displayed accordingly, a <MODIFY> ... </MODIFY> tag may be used to define the correct

order of translated words. The tags operate on small parts of a document rather than globally. To be effective the invention requires a translation program that supports Fujitsu's extended tag set. The Fujitsu invention achieves the aim of providing a machine translation 5 with higher accuracy but does so at the cost of significant pre and post processing that slows the translation. Using the Fujitsu approach it is not possible to provide machine translations in a seamless manner.

Recently granted United States patent number 6073143, 10 assigned to Sanyo Electric Co. Ltd describes a process to enhance the translation of HTML documents by adding a translation command to each hyperlink in the document. The invention seeks to address the problem of lost hyperlinks that occur during translation. It does not address improved translation of the actual document.

15

DISCLOSURE OF THE INVENTION

In one form, although it need not be the only or indeed the broadest form, the invention resides in a translation information segment associated with a communication: 20 said translation information segment including global parameters for effecting a translation of said communication or a part or parts thereof from a source language to a target language; and said translation information segment being readable to provide translation parameters to obtain the translation of the communication 25 from said first language to said target language.

The translation information segment may be embedded in the communication or attached to the communication. Alternatively the translation information segment may be stored in an accessible database and a pointer is either embedded or attached to the 30 translatable communication.

The global parameters may be selected from parameters including, but not limited to: source language, encoding, tense, available translation, translation engine, dictionary, glossary, context, translation service, individual translator, and translation memory.

- 5 The translation information segment is preferably readable by the application reading the communication, such as a web browser for web pages, an email program for email, a word processor for text documents. Alternatively, a purpose specific application may detect and action the translation information segment.
- 10 There may be two or more translation information segments associated with a communication. Each translation information segment includes parameters for translation of a portion of the communication associated with the translation information segment.

In another form the invention resides in a translation information segment associated with a communication, said translation information segment comprising at least one of: a pointer to a translation of the communication; a pointer to a translation engine for translating the communication; or a pointer to a human translator skilled in translating the communication.

- 15
- 20 The pointer to a translation of the communication is suitably a universal resource locator and there is preferably a list of pointers pointing to different language translations.

The translation information segment preferably also includes a list of translation parameters or a pointer to a file containing a list of translation parameters. The translation parameters are suitably readable by a translation engine or a human translator to improve the quality of translation.

In a still further form the invention resides in a method of providing a translated communication to a recipient of a foreign language communication including the steps of:

- 30 associating a translation information segment with the foreign

language communication;

transmitting the foreign language communication and translation information segment to a receiver;

parsing the foreign language communication to identify and analyze

5 the translation information segment; and

obtaining a translation of the foreign language communication according to parameters in the translation information segment.

When a translation is requested from a browser, the translation information segment information may be extracted from

10 the communication and forwarded to a translation manager along with a translation request.

Alternatively when a browser receives a communication to display, it may first check the translation information segment to ensure the language is correct before displaying, if not it may request

15 a translation from a translation manager.

Another alternative is for a web server to obtain a users preferred language and compare it to the translation information segment, if it does not match, then the web server could request the communication to be translated and provide the relevant details from

20 the translation information segment to the translation manager.

Yet another alternative is for the machine translation engine to view the translation information segment directly and use that information to perform a better translation.

It does not matter if the translation information segment

25 information is used at the client or the server end, the key is the information within the translation information segment is used to help the translation manager obtain the best translation.

In some case, the translation manager could be bypassed. For example if the browser views the translation information segment

30 when requesting a translation and sees a URL where the desired

translation is available, the browser could simply request that translation from the said URL.

BRIEF DESCRIPTION OF THE DRAWINGS

5 To assist in understanding the invention, preferred embodiments will be described with reference to the following figures in which:

FIG 1 shows a flowchart of seamless translation process;

FIG 2 shows a flowchart of a non-seamless translation process; and

10 FIG 3 shows a system overview of a translation process utilizing a translation information segment.

DETAILED DESCRIPTION OF THE DRAWINGS

Referring to FIG 1, there is shown a flowchart of a seamless method of translating a communication from a first language to a second language. For ease of description the method is described in respect of a single translation of a web page from a source language to a target language, it will be appreciated that it is trivial to extend the process to translate multiple communications to multiple languages. Furthermore, any communication can be translated according to the method including text documents, email, SMS messages, and audio files, video, etc.

A key element of the method is the inclusion of a translation information segment (TIS) with the communication. The TIS provides information that can make the translation seamless to sender and recipient of the communication. In the simplest form, the TIS provides one piece of information to help obtain the best translation of the communication. The one piece of information could be any of the parameters described in this application such as a URL that already

has a professionally translated version of the communication available, or a translation memory that already has many of the phrases and sentences translated.

The TIS is not just about obtaining a better machine

5 translation, but is about trying to obtain the best translation. This means leveraging off an existing human translation where possible, so if a professionally human translated version is available, then it may be obtained. Or if human translated segments are available in a translation memory, they may be obtained.

10 In a more complex form (described in detail below) the TIS includes a fuller list of all parameters for obtaining a better translation of the communication to the target language, including such parameters as tone, subject matter, preferred dictionary, preferred glossary, preferred translation engine, words to exclude, data to

15 ignore, translation service and payment method (for commercial translations). It will be realized that some of these parameters are useful to human translators as well as machine translators, for example a dictionary or glossary. The benefit of the TIS is not limited to machine translation.

20 The TIS may alternatively consist of a pointer to a file that contains all of the parameters listed above. This embodiment is essentially equivalent to having the information embedded with the communication but may be more efficient in a network environment where a translation manager maintains a database of translation

25 parameters which are retrieved according to the TIS identifier at the time of translation.

A seamless translation system may also be provided for communications that do not include a TIS but this process is described in our co-pending application titled "Seamless Translation System".

30

There may be a different TIS associated with different parts of a communication. For example, a communication may include a quoted section that has a different tone from the rest of the communication. To obtain a quality communication the TIS for the 5 bulk of the document will have a different set of parameters from the TIS associated with the quoted section. For ease of explanation a single TIS is described associated with a single communication but the invention may be extended to multiple TIS with each communication.

10 In FIG 1, the method commences when a user requests a web page. The users browser parses the web page for the TIS. An example generic TIS may have the following structure:

```

15 <?xml version="1.0" encoding="UTF-8"?>
<TIS>
  <Version>1.0</Version>
  <SourceLang>en</SourceLang>
  < MIME-Type>text/rtf</MIME-Type>
  <Encoding>ISO8859-1</Encoding>
20  <Tense>
    <Item1>formal</Item1>
    <Item2>business</Item2>
  </Tense>
  <AvailableTranslation>
25    <de_DE>http://www.source.com/reference</de_DE>
    <fr_CA>file://lanhost//d:/path/docname</fr_CA>
  </AvailableTranslation>
  <TranslationMemory>
    <Item1>TM-1 reference</Item1>
30    <Item2>TM-2 reference</Item2>
  </TranslationMemory>
  <Service>
    <Engine>special engine xyz</Engine>
    <PreferredAgency>Worldlingo</PreferredAgency>
35    <PreferredTranslator>
      <Item1>
        <Language>de</Language>
        <Name>Hans Schmidt</Name>
      </Item1>
      <Item2>
        <Language>it</Language>
40

```

```

        <Name>Bruno Zagani</Name>
        </Item2>
        </PreferredTranslator>
    </Service>
    5    <Dictionary>
        <Item1>Dictionary 1 reference</Item1>
        <Item2>Dictionary 2 reference</Item2>
    </Dictionary>
    <Glossary>
    10   <Item1>Glossary 1 reference</Item1>
    </Glossary>
    <DNT>
        <Item1>Microsoft</Item1>
        <Item2>Worldlingo</Item2>
    15   </DNT>
    <DNT-List>
        <Item1>http://www.source.com/dnt-list-doc</Item1>
    </DNT-List>
    <Use>
    20   <Item1>marketing</Item1>
    </Use>
    <Industry>
        <Item1>engineering</Item1>
    </Industry>
    25   </TIS>

```

The actual markers will vary in any given situation and may include a subset of those shown in the example or additional markers not shown.

The markers shown in the generic XML TIS have the following functions:

```

<TIS> ... </TIS>
    marks the start and end of the TIS;
<Version> ... </Version>
    indicates the version of the TIS structure;
    35 <SourceLang> ... </SourceLang>
        marks the language of the communication;
<MIME-Type> ... </MIME-Type>
    indicates the MIME type;
<Encoding> ... </Encoding>
    40   indicates the encoding;
<Tense> ... </Tense>
    indicates the tense, this is read by the machine translation
    engine as a parameter that may improve the quality of the
    translation;
    45 <Item#> ... </Item#>

```

delimits multiple items for indicating priority. Item 1 applies before item 2;

5 <AvailableTranslation> ... </AvailableTranslation>
 lists available/preferred translations. For example, a web page may already have a foreign language equivalent that can be delivered instead of the accessed page;

10 <TranslationMemory> ... </TranslationMemory>
 points to a translation memory for retrieval of translations from a cache to avoid retranslation of translated documents or parts of documents;

15 <Service> ... </Service>
 indicates preferences for the translation such as a particular translation engine or particular human translators;

20 <Engine> ... </Engine>
 the preferred engine;

25 <PreferredAgency> ... </PreferredAgency>
 the preferred agency for performing required translations;

30 <PreferredTranslator> ... </PreferredTranslator>
 the preferred human translator, perhaps according to each language;

35 <Dictionary> ... </Dictionary>
 the translation dictionary or dictionaries to be used;

40 <Glossary> ... </Glossary>
 the translation glossary or glossaries to be used;

45 <DNT> ... </DNT>
 a list of words or phrases not to translate;

50 <DNT-List> ... </DNT-List>
 a pointer to a file containing a list of words or phrases not to translate;

55 <Use> ... </Use>
 a context marker used by the translation engine to improve the quality of translation;

60 <Industry> ... </Industry>
 another context marker for improving the quality of translation.

65 Once the TIS is identified the browser extracts the translation parameters and performs actions accordingly. The first action is to check the source language tag against the preferred language of the user. The preferred language of the user may be obtained from the operating system setup, a cookie, a preferences file residing on the recipients computer or other accessible location, or from an analysis performed by suitable software. If the preferred language matches the source language there is no translation necessary and the page is displayed. If there is not a match a translation is obtained.

The TIS is not limited to facilitating a seamless translation of a communication. The TIS will also improve the quality of translation in a non-seamless system, such as the one-click translation system described in our co-pending application mentioned earlier. A non-
5 seamless translation system utilizing the TIS is shown in FIG 2.

The process depicted in FIG 2 commences when a user receives, for example, an email and the email program displays the email. If the email is not in the preferred language of the recipient an action, such as clicking a one-click translation component, is taken to
10 request a translation. A translation manager parses the email for a TIS. The content of the TIS is analyzed and a translation obtained according to the information contained in the TIS. For email this will normally mean supplying translation parameters to translation engine.

15 An advantage of the TIS is that it may contain a redirection to a foreign equivalent of a requested communication. Many businesses maintain mirror sites in multiple languages. The TIS may contain pointers to these sites, as indicated in the previous generic sample.

As mentioned above, the TIS is not limited to web pages. The
20 TIS can be added to email in a similar manner to the known use of VCARDS. The following example shows the attachment of a TIS to an email using a custom MIME-type (also called Content-Type) such as "text/x-tis".

25 Content-Type: text/x-tis; name="settings.tis"
Content-Transfer-Encoding: quoted-printable
Content-Disposition: attachment; filename="settings.tis"
<TIS>
...
</TIS>

30 A TIS, such as shown in the earlier example, is embedded into a separate part of an email. These separate parts inside emails are common practice and represent attachments to the given email content. An advantage of the TIS as an attachment is that it is

unaffected by transmission across the internet and is not dependent upon the mail handling system of individual mail servers.

An alternative embodiment for implementing the TIS in email is to add a custom header of the following form:

5 Received: from Laptop (isp.net [192.168.41.217]) by wlm.worldlingo.com
 with SMTP (Microsoft Exchange Internet Mail Service Version #)
 id 1BV3QHTA; Fri, 9 Feb 2001 09:29:49 +1000
From: "###" <#worldlingo.com>
10 To: "###" <#worldlingo.com>
 Subject: Patent
 Date: Thu, ## ### ##
 Message-ID: <#####.#@worldlingo.com>
 MIME-Version: 1.0
15 Content-Type: text/plain;
 charset="iso-8859-1"
 Content-Transfer-Encoding: 7bit
 X-Priority: 1 (Highest)
 X-MSMail-Priority: High
20 X-Mailer: Microsoft Outlook IMO, Build 9.0.2416 (9.0.2911.0)
 X-MimeOLE: Produced By Microsoft MimeOLE V5.50.4133.2400
 X-TIS-Version: 1.0
 X-TIS-SourceLang: en
 X-TIS-Service: Engine=engine1
25 X-TIS-Tense: formal, business
 Importance: High

This example shows only a few fields of the possible TIS fields noted earlier in the generic XML structure example. Only required fields are included in the TIS for any particular application. Fields that 30 are not required may be replaced with well-known and reasonable default values, or simply omitted.

The header embodiment of the TIS may also be applied to documents in text, RTF, or proprietary formats. All documents contain header information that dictates the appearance of the 35 document. The TIS can be added to this header information so that the document is seamlessly translated before being viewed by the receiver. The TIS could also be added to the properties dialog box of a document created using MSWord® or other proprietary word processors.

The TIS can also be included as part of HTML documents as shown in the following example of an HTML comment block.

```

5 <html xmlns:t="urn:schemas-worldlingo-com:tis:tis"
      xmlns="http://www.w3.org/TR/REC-html40">
<head>
<meta name="X-TIS-Version" content="1.0">
<title>Reference Document</title>
<!--><xml>
10 <t:TIS>
    <t:SourceLang>en</t:SourceLang>
    <t:Version>1.0</t:Version>
    <t:MIME-Type>text/rtf</t:MIME-Type>
    <t:Encoding>ISO8859-1</t:Encoding>
15 <t:Tense>
    <t:Item1>formal</t:Item1>
    <t:Item2>business</t:Item2>
</t:Tense>
<t:Industry>
20 <t:Item1>engineering</t:Item1>
</t:Industry>
</t:TIS>
</xml><!-->

```

An alternative solution for using the TIS with HTML documents
25 as shown in the following example of an HTML meta tag.

```

<HTML>
<HEAD>
<META name="X-TIS-Version" content="1.0">
30 <META name="X-TIS-SourceLang" content="en">
<META name="X-TIS-Service" content="Engine|engine1">
<META name="X-TIS-Tense" content="formal,business">
<TITLE>Search Results</TITLE>
</HEAD>
35 <BODY>
...
</BODY>

```

The TIS is not limited to text applications. Rudimentary
translation engines are available for translating voice to text, text to
40 voice, and voice to voice. The TIS can dramatically improve the
usefulness of these rudimentary translation engines by defining
parameters such as tone, accent, content and field.

A schematic of a practical implementation of the TIS in a network environment is shown in FIG 3. A user 1 requests or receives a communication, such as a web page 2, using a browser on a personal computer 3. The browser requests the page 2 from a 5 web server 4 via the internet 5, and it is displayed on the personal computer 3.

If the communication is in a language foreign to the user, the user 1 may request a translation. As discussed above, this step may occur automatically according to the process described in our co-pending application. The browser on the personal computer parses 10 the communication for a TIS and requests a translation via the internet 5 using the parameters obtained from the TIS.

If the web server 4 has a suitable translation 2a of the communication 2 it is supplied directly to the user 1. If a suitable 15 translation is not available the translation request is passed to a translation manager 6 with the parameters from the TIS. The translation manager 6 obtains the translation 2b from a translation engine 7.

For ease of explanation the translation manager 6 and 20 translation engine 7 have been shown separately. These functions may be embodied in a single application or separate applications running on a single computer. The translation functions may even be performed locally on the personal computer 3 if appropriate software is installed.

25 The TIS may be read by the application receiving the communication but is not limited to this implementation. If a translation engine is resident on the receiver's computer, or in a network to which the receiver is connected, the translation engine may directly interpret the TIS. More suitably, a server in the network 30 may be configured as a translation manager that detects a TIS and manages the translation of the communication before delivering the

communication to the recipient. The translation manager may be resident on the computer of the recipient.

The application of the TIS to specific cases will now be explained to assist with understanding the invention.

5 Most existing on-line businesses have originated in the west and have developed their web pages and documents in English. However, the fastest growing Internet access is occurring in areas where English is not the first language and may not even be spoken by many people who may be potential customers. In order to market

10 to these potential customers a web page and documents must be presented in their native language. Most people will not go to the trouble of translating a page and certainly will not pay for the translation. In order to market to these people the translation must occur seamlessly.

15 The translation may occur in a number of ways, all of which are facilitated by the TIS. Firstly, the recipient may have machine translation software resident on their computer. In this case the TIS provides all relevant parameters to seamlessly result in display of a high quality translation. Secondly, the recipient will be attached to the

20 Internet so the TIS can direct the web page to a translation manager that makes the necessary translation and displays it seamlessly to the recipient. Thirdly, the originator may have already produced a mirror site in the relevant language, in which case the TIS seamlessly directs the browser of the recipient to the mirror site.

25 To achieve maximum effectiveness the parameters contained in the TIS must be relevant and understandable by the translation engine being employed. As there is a wide range of translation engines this requirement could present difficulty. However, the inventor has realized that the TIS contains an extendible generic set

30 of parameters. It is a relatively straightforward problem for a machine translation engine to interpret the TIS and convert the generic

parameters into specific commands. The inventor envisages that it would also be possible to generate conversion programs to interpret the generic TIS parameters for legacy translation engines.

Throughout the specification the aim has been to describe 5 embodiments of the invention without limiting the invention to any specific combination of alternate features.

Dated this 22nd Day February of 2001

WORLTLINGO Inc.

10 By their Patent Attorneys

FISHER ADAMS KELLY

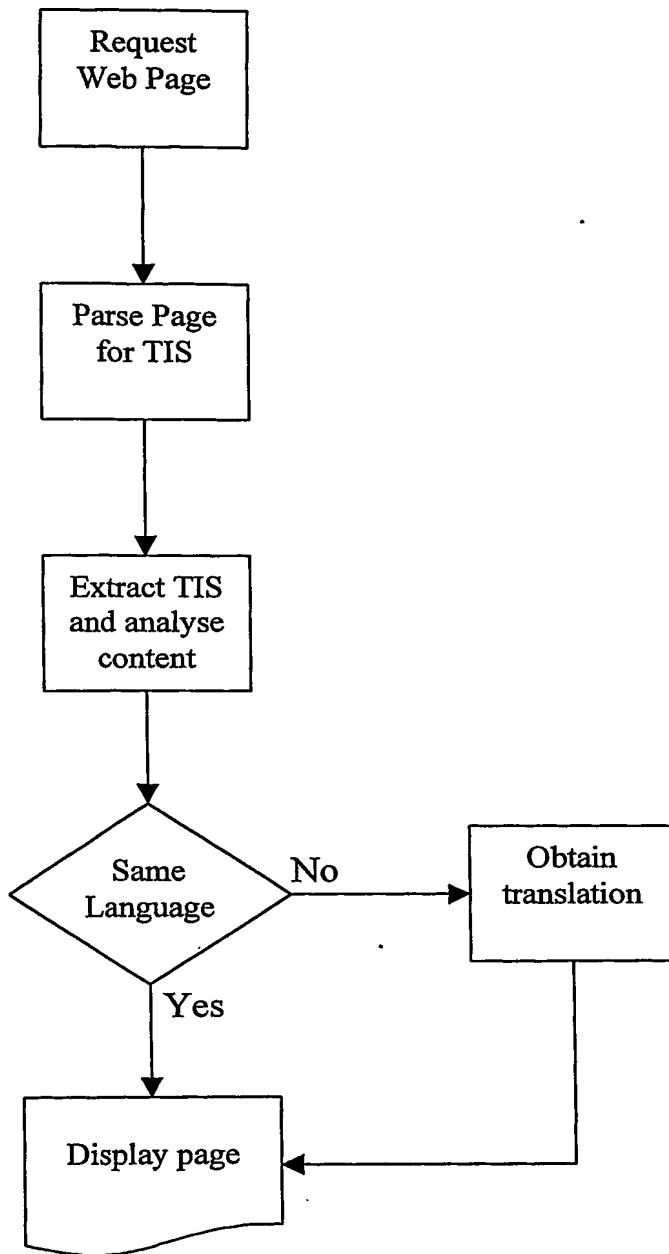


Fig 1

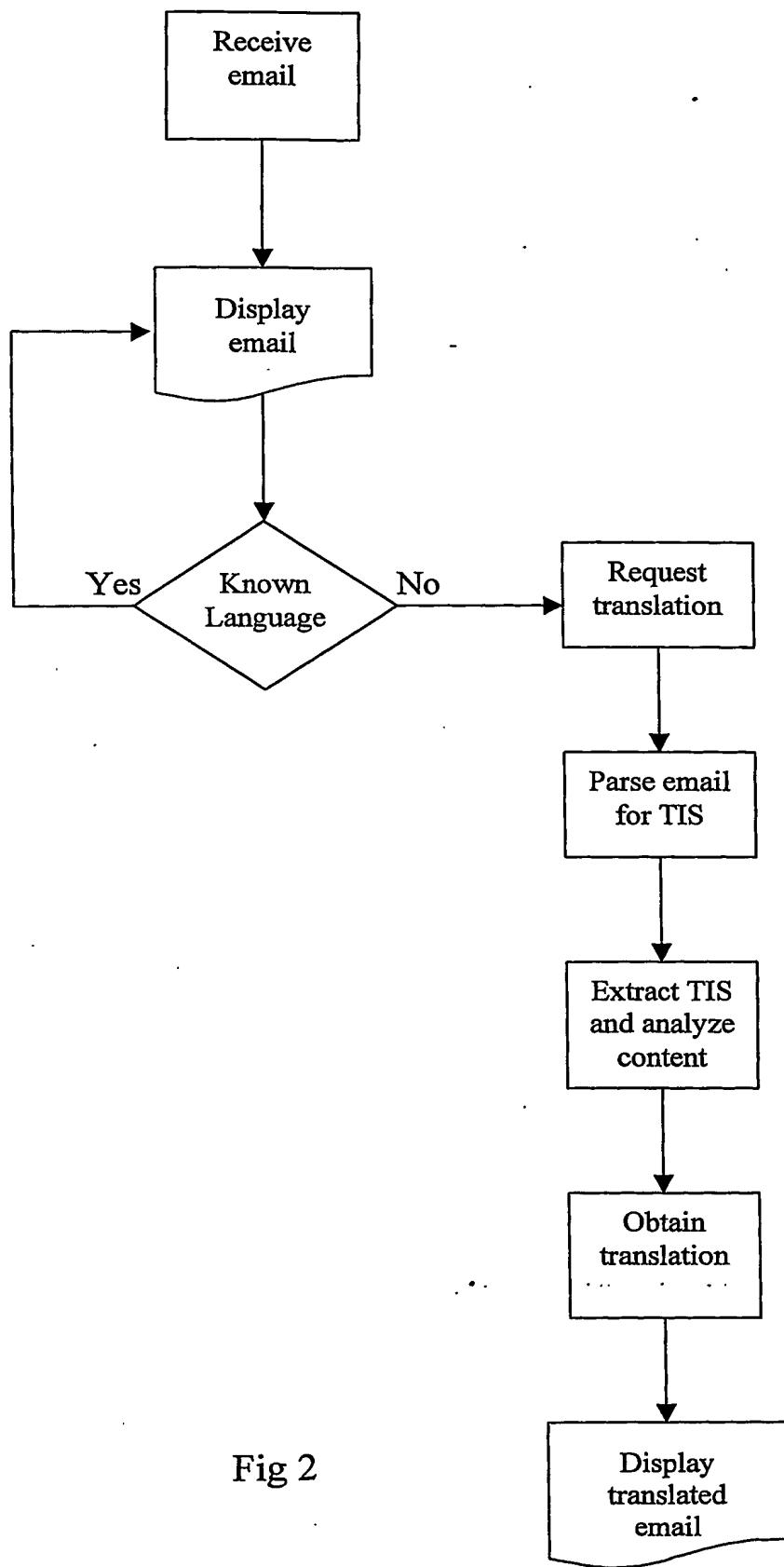


Fig 2

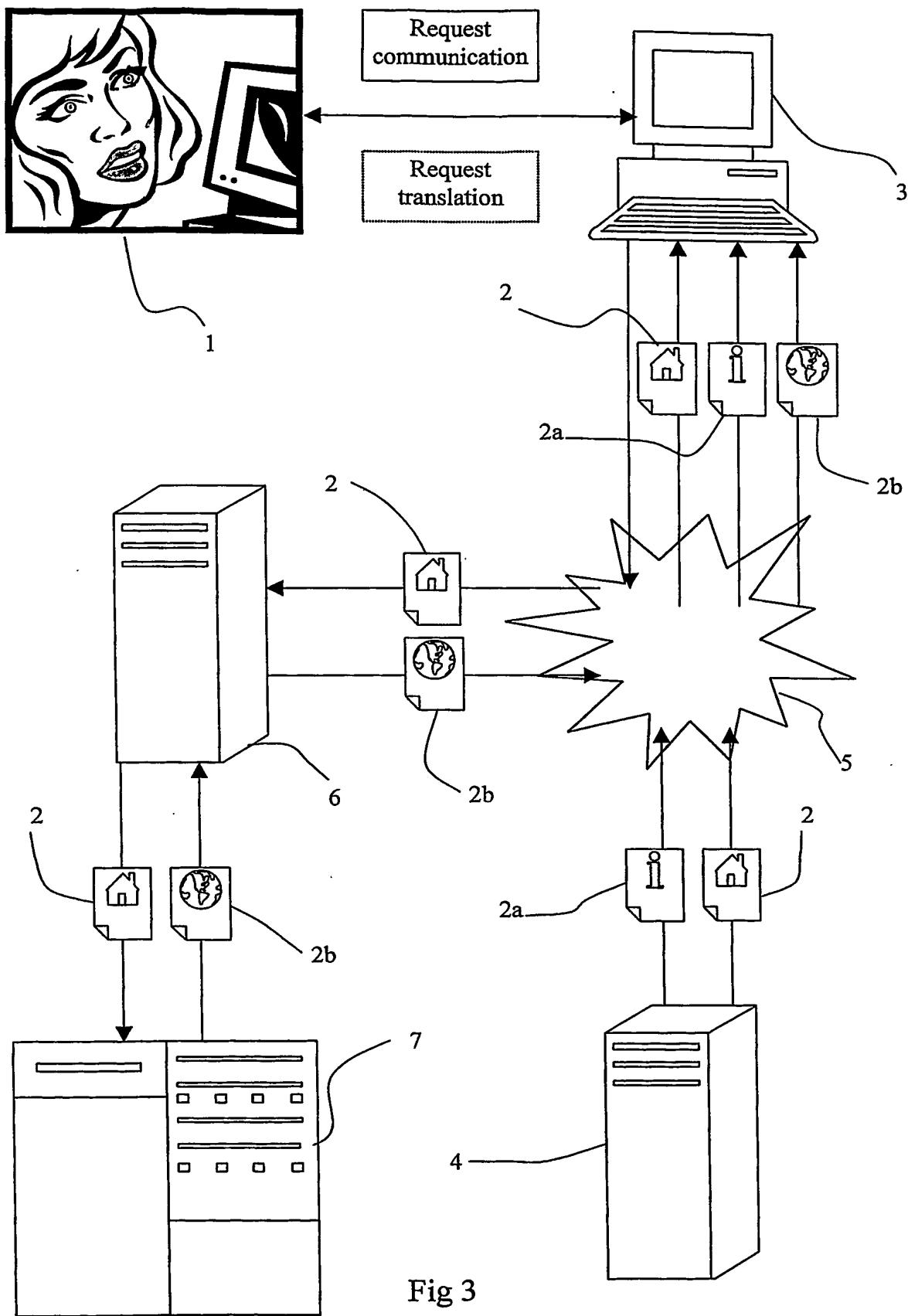


Fig 3